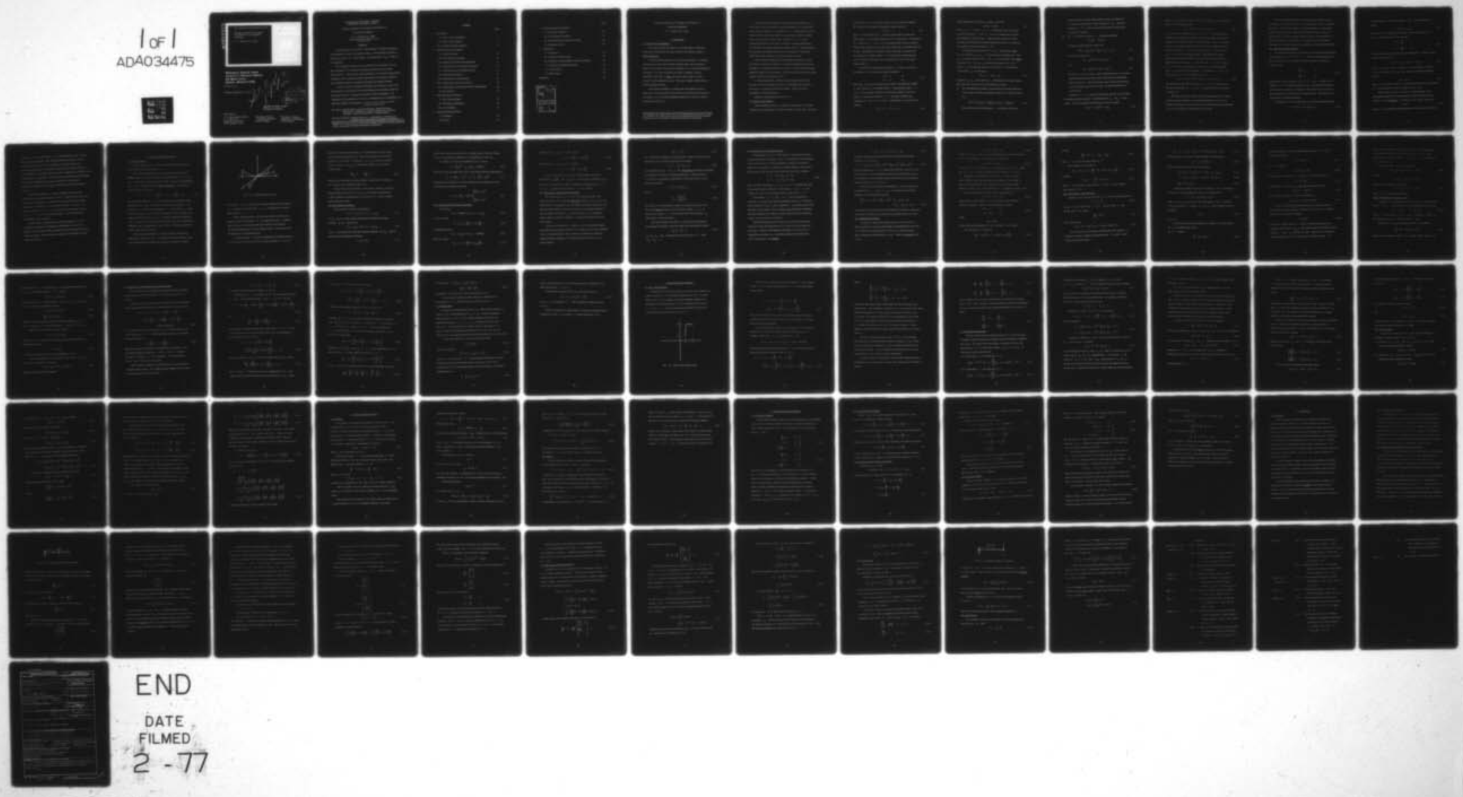
AD-A034 475 WISCONSIN UNIV MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
GENERAL ESTIMATES FOR LINEAR FUNCTIONALS IN NONLINEAR PROBLEMS.(U)
DEC 76 M J SEWELL, B NOBLE DAAG29-75-C-0024
UNCLASSIFIED MRC-TSR-1703 NL

1 OF 1
ADA034475



END
DATE
FILMED
2-77

MRC Technical Summary Report # 1703

GENERAL ESTIMATES FOR LINEAR FUNCTIONALS IN NONLINEAR PROBLEMS

M. J. Sewell and B. Noble

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

December 1976

(Received September 29, 1976)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

University of Oxford
Computing Laboratory
Oxford, England

University of Reading
Department of Mathematics
United Kingdom

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

GENERAL ESTIMATES FOR LINEAR FUNCTIONALS IN NONLINEAR PROBLEMS

M. J. Sewell and B. Noble

Technical Summary Report # 1703
December 1976

ABSTRACT

A general theory is developed for the estimation of linear functionals, in three distinct classes of nonlinear problems. The functional is linear in the solution vector x_0 of the problem, an example being (x_0, p) where p is assignable.

The considered problems are all generated via the gradients of some given quadratic or non-quadratic Lagrangian functional over two inner product spaces. This may be a saddle functional, or it may be constructed by embedding a given nonlinear problem with the aid of a Lagrange multiplier. Many different problems in applied mathematics are thereby included.

In some cases the assignable coefficient can be chosen in such a way that the bounds calculated for the linear functional are pointwise bounds on the solution vector. In general this requires further investigation, but estimation of the deflection at a point on a cantilever beam is illustrated in §6.

AMS(MOS) Subject Classifications: 35-A15, 47-H15, 49-D45, 49-H05, 65-N30, 69.49, 73.49

Key Words: Approximation of linear functionals, Saddle functional, Optimization, embedding, nonlinear programming, Extremum principles, variational principles, Pointwise bounds

Work Unit Numbers 1 (Applied Analysis), 3 (Applications of Mathematics), 5 (Mathematical Programming and Operations Research)

This research was sponsored in part by the United States Army under Contract No. DAAG29-75-C-0024, and in part by the University of Oxford Computing Laboratory and the University of Reading Mathematics Department.

Contents

Page

GENERAL ESTIMATES FOR LINEAR FUNCTIONALS IN

NONLINEAR PROBLEMS

M. J. Sewell and B. Noble

1. Introduction

(i) Scope of the investigation

This paper presents the results of a general study of systematic methods for getting upper and lower bounds to the solution-values of linear functionals.

A new theoretical framework is set up for this purpose. It contains a wide class of linear and nonlinear problems which can be defined in terms of the gradients of some given quadratic or non-quadratic generating functional. It is often important to be able to construct, using an assignable coefficient, a linear functional of the solution of such a problem, and to estimate its value. This can be related to the problem of finding pointwise bounds.

The framework exhibits in a natural way three different types of situations, requiring different methods which we call the general optimization method (§ 2), the general embedding method (§ 4), and the nonlinear programming method (§ 5).

This research was sponsored in part by the United States Army under Contract No. DAAG29-75-C-0024, and in part by the University of Oxford Computing Laboratory and the University of Reading Mathematics Department.

Underlying these situations is an appropriate generalization (§ 3) of the second mean value theorem, already indicated in our earlier joint paper (Noble and Sewell, 1972, equation (5.3)). In particular this can be used to provide sufficient conditions for satisfying a saddle inequality in the form proposed by Sewell (1969, equation (2.50)). A saddle functional generates a wide class of problems in applied mathematics, as described in the papers cited and in Sewell, 1973a, b, where elasticity and plasticity are treated in detail from this viewpoint. The general optimization method applies to saddle-generated problems.

Under different hypotheses on the generating functional, such as boundedness (instead of positivity) of operators representing its second derivatives, the saddle hypothesis may be lost. In this case the general embedding method can be available. We show how it recovers some recent results of Barnsley and Robinson (1976).

Problems generated by inserting a given scalar functional into governing conditions expressed as sets of inequalities are covered in the section on nonlinear programming methods. They also lead to inequalities on linear functionals.

Remarks on applications are made in §6.

(ii) Origin of the research

This investigation began in an attempt to generalize to nonlinear problems some approximation methods described by Fujita (1955), who gave

an elementary proof of a theorem of Kato (1953) on pointwise estimates for a solution of the linear decomposable operator equation

$$T^* T x = f . \tag{1.1}$$

Here x is the unknown and f is given, both 'vectors' in the same linear space. T is a linear operator and T^* is its adjoint. For example, if $T \sim$ grad and $T^* \sim -$ div, (1.1) is associated with Poisson's equation. Fujita's paper subsumes in a compact way earlier work on pointwise bounds by Diaz, Greenberg and Weinstein, Prager and Synge, and others (see the references in Fujita's paper). It is convenient to recapitulate here some of Fujita's conclusions, as an introduction to some of the ideas required later on.

We introduce an intermediate variable u in order to decompose the problem (1.1) into the pair of operator equations

$$\left.\begin{array}{lr} T^* u = -f , & (\alpha) \\ T x = -u . & (\beta) \end{array}\right\} \tag{1.2}$$

Both here, and in the main general theory below, we regard the variable x as an element of a real vector space E having inner product $(\cdot,\cdot)$, and u as an element of another, and normally different, real vector space F having inner product $\langle\cdot,\cdot\rangle$. The linear operators map subspaces E' and F' of E and F (respectively) according to the scheme

$$T : E' \to F, \qquad T^* : F' \to E . \tag{1.3}$$

Mutual adjointness of these two operators means that

$$(x, T^* u) = \langle u, Tx \rangle \tag{1.4}$$

for all x in E' and all u in F'. For example, when differential operators form part of T and T^*, (1.4) is a compact way of writing the integration by parts formula. Many examples of these and other relevant simple ideas from functional analysis are given in an Appendix to the paper of Noble and Sewell (op. cit.).

We emphasize those values of x and u which satisfy both (1.2α) and (1.2β) by x_0, u_0, i.e. by attaching a subscript zero. Thus x_0 is an actual solution of (1.1). Let u_α be any solution of the single constraint (1.2α). Let x_β, u_β be any pair satisfying *only* (1.2β), so that x_β is an arbitrary vector in the domain of T and generates a consequent u_β. In other words

$$T^* u_\alpha = -f, \qquad Tx_\beta = -u_\beta . \tag{1.5}$$

In general $u_\beta \neq u_\alpha$ unless both are u_0 belonging to the actual solution. Then Fujita's conclusions can be summarized as follows.

(a) The dual extremum principles, giving what can be called upper and lower "energy" bounds in appropriate contexts, are (Fujita, equation (2.3))

$$\frac{1}{2} \|u_\alpha\|^2 \geq \frac{1}{2} (x_0, f) = \frac{1}{2} \|Tx_0\|^2 \geq (x_\beta, f) - \frac{1}{2} \|Tx_\beta\|^2 . \tag{1.6}$$

The norms here are all in the space F, but later on it will not

cause confusion to use the same symbol for norms of elements of E. These principles bound a linear functional of x_0, with given coefficient f in E. This can be the actual work of given forces in mechanical problems.

(b) If q is an arbitrary vector in F', the linear functional

$$\langle u_0, q\rangle = -(x_0, T^* q)$$

is bounded on both sides by (Fujita (3.6))

$$\frac{1}{2}\|u_\alpha - u_\beta\| \, \|q\| \geq |\langle \frac{1}{2}(u_\alpha + u_\beta) - u_0, q\rangle| \tag{1.7}$$

and also by (Fujita (3.8))

$$\|u_\alpha - u_\beta\| \, \|q\| \geq |\langle u_\beta - u_0, q\rangle| \tag{1.8}$$

and

$$\|u_\alpha - u_\beta\| \, \|q\| \geq |\langle u_\alpha - u_0, q\rangle| \, . \tag{1.9}$$

We call (1.8) and (1.9) Fujita's 'weak' estimates and (1.7) his 'strong' estimate because more is given away to get the weak inequalities than the strong one. We shall recover some of these results below, by proofs different from those of Fujita, as simple illustrations of our framework.

Equations (1.7) - (1.9) suggest the following approach to the problem of obtaining pointwise bounds. Remembering that $u_0 = -Tx_0$, choose q so that $T^* q$ has a delta function behavior in such a way that

$$\langle u_0, q\rangle = -(x_0, T^* q) = -(x_0)_p \tag{1.10}$$

where $(x_0)_P$ denotes the value of the exact solution x_0 at the point P. Then (1.9), for example, gives

$$\|u_\alpha - u_\beta\| \|q\| \geq |(x_0)_P + \langle u_\alpha, q\rangle| , \tag{1.11}$$

and we have found pointwise bounds on x_0 at P. This procedure is useful for one-dimensional problems and in two and three dimensional problems for bounding quantities on the boundaries. But if, for instance, we try to bound the potential at an interior point in a problem involving Poisson's equation, q has to behave like grad$(1/r)$ near this point and $\|q\|$ involves a divergent integral. This difficulty has been circumvented by various authors in an ingenious way, the essence of which depends on choosing q to have the form $q = q' - Tp'$, where T^*q' and T^*Tp' have exactly the same type of δ-function behavior, with q' such that (1.10) is true with q' in place of q, and p' is in the domain of T. We can deduce from (1.9)

$$\|u_\alpha - u_\beta\| \|q' - Tp'\| \geq |\langle u_\beta, q\rangle - (f, p') + (x_0)_P| . \tag{1.12}$$

The expression on the left is finite since q' and Tp' are chosen so that their singularities at P cancel. A numerical example is discussed in Fujita (1955).

Although we have been able to obtain pointwise bounds in a number concrete nonlinear problems by essentially generalizing (1.11), as for example in §6 below, we have not been able to find a natural generalization of (1.12) in the abstract nonlinear setting of our work.

Our formalism throws light on some bounding principles developed by Martin (see, for example 1964, 1966) in the special context of elasticity and creep. Martin exploits ideas connected with energy, complementary energy, virtual work, etc. We show in § 6 that his formulae apply in a quite general context by exploiting simply convexity and the structure of the basic equations. The relevance of convexity in the dual extremum principles of continuum mechanics was originally pointed out by Hill (1956).

(iii) General governing equations

Our general theory is set in the same two inner product spaces having typical elements x in E and u in F which are described after (1.2). We consider the class of possibly nonlinear problems of generalized Lagrangian type

$$\frac{\partial L}{\partial x} = 0 \, , \qquad (\alpha) \qquad\qquad (1.13)$$

$$\frac{\partial L}{\partial u} = 0 \, , \qquad (\beta)$$

generated by a given functional $L[x,u]$ of x and u (Sewell, 1973a, equations (25)). The partial gradients in (1.13) are Gateaux differentials, as in the familiar process of computing a 'first variation' and picking out the coefficients of increments in the varied argument. Thus the pair (1.13) is effecting a variational principle.

The problem (1.2) is recovered from (1.13) with the special example

$$L = (x, T^{*}u) + (f,x) + \frac{1}{2}\langle u,u\rangle \, , \qquad (1.14)$$

using adjointness (1.4) before computing $\partial L/\partial u$. More generally the example

$$L = (x, T^* u) - X[x,u] , \tag{1.15}$$

where $X[x,u]$ is another given possibly nonquadratic functional of x and u, generates from (1.13) the equations

$$\begin{aligned} T^* u &= \frac{\partial X}{\partial x} , && (\alpha) \\ Tx &= \frac{\partial X}{\partial u} , && (\beta) \end{aligned} \tag{1.16}$$

of Hamiltonian type proposed for study by Noble (1964). Another concrete example is generated by

$$L = (x, T^* u) + (f, x) - \frac{1}{2}\rho(x,x) + \frac{1}{2}\langle u,u\rangle \tag{1.17}$$

where ρ is a given scalar. As with (1.2), it is possible to eliminate u from (1.13) with (1.17) and recover a single decomposable generating equation

$$(T^* T + \rho I)x = f \tag{1.18}$$

where I is the identity operator. Dual extremum principles for this when $\rho \geq 0$ were studied by Noble and Sewell (op. cit., §14).

Problems whose ab initio version is nondecomposable via an intermediate variable in the above sense may still be brought into the scheme (1.13) by <u>embedding</u>. For example, if the ab initio equation is

$$N(x) = 0 \tag{1.19}$$

where N is a possibly nonlinear operator, we may seek to identify this

equation as (1.13β) by introducing a u to appear linearly in some $L[x,u]$ like a Lagrange multiplier. This embedding procedure induces a second 'adjoint' equation (1.13α) to be considered in conjunction with (1.19), and perhaps containing an assignable coefficient in the linear functional to be estimated. It is in this way that the work of Barnsley and Robinson (1976) is brought into our framework. Even if the ab initio problem is decomposable, it may still be embedded in the stated manner into a larger problem. Barnsley and Robinson (1974) do this in their study of the linear equation (1.18) for $\rho = 1$.

In the general problem (1.13) certain additional hypotheses are required about the general functional $L[x,u]$. Typically these set bounds on the second derivatives. In particular the functional may be a saddle functional. For example, (1.17) is strictly convex in u and, if $\rho > 0$, strictly concave in x. If $\rho = 0$ as in (1.14) it is only weakly concave in x. If $\rho < 0$ it is not a saddle functional. Such hypotheses are made precise in the next Section.

Another source of generalization is that the equations (1.13) can be replaced by systems of inequalities (inequalities (33) and (34) of Sewell, 1973a) and dual extremum principles can still be proved under the saddle hypothesis. These sometimes contain direct estimates for linear functionals (see § 5).

2. General Optimization Method

(i) Saddle functional

Suppose that $L[x,u]$ is a given saddle functional defined over some domain in the product space $E \times F$.

The analytical expression of the saddle property is in terms of arbitrary pairs of 'points' in this domain, which we label x_+, u_+ and x_-, u_- and refer to as the 'plus point' and the 'minus point' respectively. Then $L[x,u]$ is called a (weak) saddle functional if, for any pair of distinct points in its domain,

$$L_+ - L_- - \left(x_+ - x_-, \left.\frac{\partial L}{\partial x}\right|_+\right) - \left\langle u_+ - u_-, \left.\frac{\partial L}{\partial u}\right|_-\right\rangle \geq 0 . \tag{2.1}$$

The subscripts attached to L and its gradients mean evaluation at the indicated points. Such a functional is concave with respect to x at each fixed u, and convex with respect to u at each fixed x - hence the name, and Fig. 2.1 is a schematic illustration of its individual cross-sections with the spaces E and F. The weak inequality permitted in (2.1) for distinct pairs of points means that the surface can contain linear segments such as straight lines or plane facets. Otherwise it would be called a strict saddle functional.

This analytical statement of a saddle functional was given by Sewell (1969, equation (2.50)). For simplicity in what follows we adopt the convention that the vertical bar attached to gradients is omitted.

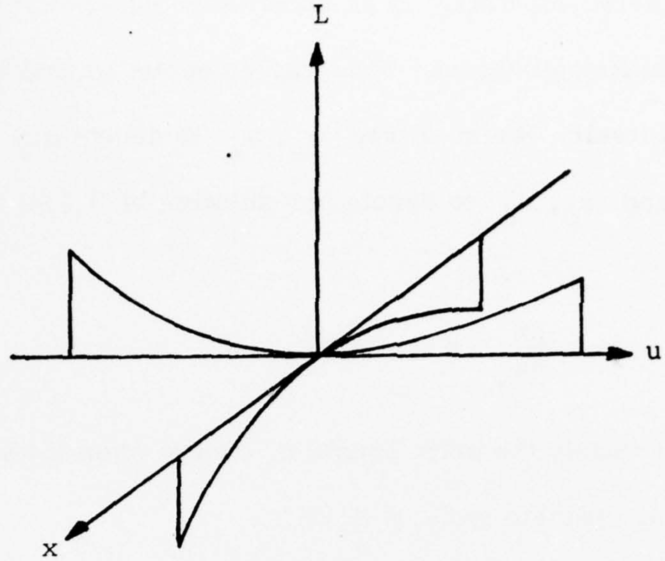


Fig. 2.1. Saddle functional $L[x,u]$

For example, $\partial L/\partial x_+$ will denote the Gateaux differential with respect to x evaluated at the plus point x_+, u_+ (and not merely a gradient with respect to x_+).

Unless otherwise stated, plus and minus points will always be arbitrary points in the domain of L, throughout the paper. Our entire theory will rest on the facility with which different and especially convenient interpretations may be assigned to them. Such choices will be indicated by an appropriate suffix.

For example, when L is used to generate the governing equations (1.13), we can divide them into the two subsets labelled (α) and (β).

Each subset considered separately is an underdetermined problem whose solutions will be supposed known. They may be easier to find than the solutions to (1.13) itself. We shall use x_α, u_α to denote any solution of (1.13α) alone, and x_β, u_β to denote any solution of (1.13β) alone. In other words,

$$\frac{\partial L}{\partial x_\alpha} = 0, \qquad \frac{\partial L}{\partial u_\beta} = 0 . \tag{2.2}$$

Neither point need satisfy the other equation, except when it happens to be a solution of the complete problem (1.13).

In what follows we shall often, for the sake of emphasis, denote an actual solution point of (1.13) by x_0, u_0, and attach a subscript zero to other quantities evaluated there, as we did for (1.2). Such a solution point need not be unique.

(ii) Dual extremum principles

First choose the particular interpretations

$$x_+, u_+ = x_\alpha, u_\alpha \quad \text{and} \quad x_-, u_- = x_0, u_0 \tag{2.3}$$

in (2.1). By (2.2) there follows immediately the *stationary minimum* principle $L_\alpha \geq L_0$. Next choose

$$x_+, u_+ = x_0, u_0 \quad \text{and} \quad x_-, u_- = x_\beta, u_\beta , \tag{2.4}$$

in (2.1). This implies the stationary maximum principle $L_0 \geq L_\beta$. Thus we arrive at the dual extremum principles

$$L_\alpha \geq L_0 \geq L_\beta \tag{2.5}$$

derived with increasing generality in our earlier papers (Noble and Sewell, op. cit., inequalities (1); Sewell 1973a, inequalities (31) and (32)).

For problem (1.1) they are illustrated by (1.6) in which

$$L_\alpha = \frac{1}{2}\|u_\alpha\|^2, \quad L_\beta = (x_\beta, f) - \frac{1}{2}\|Tx_\beta\|^2 . \tag{2.6}$$

The second order quantities given away to get these particular estimates are

$$L_\alpha - L_0 = \frac{1}{2}\|u_\alpha - u_0\|^2, \quad L_0 - L_\beta = \frac{1}{2}\|u_0 - u_\beta\|^2 . \tag{2.7}$$

The extremum principles (1.6) can be rewritten as error estimates in terms of the difference between the bounds

$$L_\alpha - L_\beta = \frac{1}{2}\|u_\alpha - u_\beta\|^2 \geq \begin{cases} \frac{1}{2}\|u_\alpha - u_0\|^2 \\ \\ \frac{1}{2}\|u_\beta - u_0\|^2 \end{cases} . \tag{2.8}$$

(iii) General bounds for linear functionals

If we choose

$$x_+, u_+ \text{ arbitrary, and } x_-, u_- = x_0, u_0 \tag{2.9}$$

in (2.1), we find

$$L_+ - L_0 - (x_+, \frac{\partial L}{\partial x_+}) \geq - (x_0, \frac{\partial L}{\partial x_+}) . \tag{2.10}$$

If instead we choose

$$x_+, u_+ = x_0, u_0 \text{ and } x_-, u_- \text{ arbitrary,} \tag{2.11}$$

then (2.1) gives

$$L_0 - L_- + \langle u_-, \frac{\partial L}{\partial u_-} \rangle \geq \langle u_0, \frac{\partial L}{\partial u_-} \rangle . \tag{2.12}$$

Next we add $L_0 - L_\beta \geq 0$ to (2.10), giving

$$L_+ - L_\beta - (x_+, \frac{\partial L}{\partial x_+}) \geq - (x_0, \frac{\partial L}{\partial x_+}) . \tag{2.13}$$

Also we add $L_\alpha - L_0 \geq 0$ to (2.12), giving

$$L_\alpha - L_- + \langle u_-, \frac{\partial L}{\partial u_-} \rangle \geq \langle u_0, \frac{\partial L}{\partial u_-} \rangle . \tag{2.14}$$

It can be seen that (2.13) and (2.14) offer bounds on the linear functionals $(x_0, \frac{\partial L}{\partial x_+})$ and $\langle u_0, \frac{\partial L}{\partial u_-} \rangle$ of the solution variables x_0, u_0. The bounds on the left are in terms of arbitrary assignable points x_+, u_+ or x_-, u_-, and the supposedly known α- and β-points.

(iv) Optimization of the extremum principles

The choices made in (2.3) and (2.4) are special choices of the pairs of points in (2.1), made with particular solutions of (2.2α) and (2.2β) respectively, and designed to lead immediately to simple conclusions (2.5). Such particular solutions need not be unique, and in specific problems it may be possible to decrease L_α and/or increase L_β by optimizing within subsets of particular solutions. In general the problem is to find such subsets.

In the case of problem (1.2), Fujita (op. cit. § 4) specifies subsets appropriate for improving the bounds (2.6). Here we make a rather different remark about that problem to help motivate our subsequent procedures. Noticing that the Lagrangian (1.14) implies that the left side of (2.1) is exactly equal to

$$\frac{1}{2}\|u_+ - u_-\|^2 , \tag{2.15}$$

we consider the possibility of minimizing this 'square of the error' among those plus and minus points which have the property

$$u_+ - u_- = \lambda u_\alpha + \mu u_\beta - u_0 \tag{2.16}$$

for disposable scalars λ and μ. This simultaneous procedure corresponds to finding the minimum of an elliptic paraboloid. A special feature of (1.2) leads the simultaneous optimization to the improved pair of dual extremum principles

$$L_\alpha = \bar{L}_\alpha \geq L_0 \geq \bar{L}_\beta \geq L_\beta \tag{2.17}$$

where

$$\bar{L}_\beta = \frac{(x_\beta, f)^2}{2\|Tx_\beta\|^2} . \tag{2.18}$$

The result of the simultaneous procedure is therefore the same as the two usual separate choices of first (trivially) setting $\lambda = 1$, $\mu = 0$, and secondly setting $\lambda = 0$ and optimizing the nonhomogeneous L_β with respect to the scale factor μ.

The special feature of problem (1.2) which leads simultaneous and separate procedures to the same result is an orthogonality property

$$\langle u_\beta, u_\alpha - u_0 \rangle = 0 , \tag{2.19}$$

i.e. any $u_\beta = -Tx_\beta$ is orthogonal to the null space of T^*, since $T^* u_\alpha = T^* u_0 = -f$.

(v) A general class of competing vectors

Looking ahead from (2.16) to the problem of optimizing the bounds for linear functionals for general $L[x,u]$, and noting that it is desirable to free Fujita's proofs from their dependence on Schwarz's inequality (which is a consequence of minimizing a single quadratic), we propose to study choices of the plus and minus points which have the properties

$$\begin{aligned} x_+ - x_- &= rx_\alpha + sx_\beta + hp + ix_0 , \\ u_+ - u_- &= \lambda u_\alpha + \mu u_\beta + kq + ju_0 . \end{aligned} \qquad (2.20)$$

Here the eight coefficients $r, s, h, i, \lambda, \mu, k, j$ are disposable real scalars which will be normalized here by taking $i = \pm 1$, $j = \pm 1$. The choices (2.3) and (2.4) are special cases of (2.20) with $h = k = 0$.

The elements p in E and q in F are to be regarded as assignable. Note that dual extremum principles such as (1.6) estimate a linear functional $\frac{1}{2}(x_0, f)$ whose coefficient f was already given in the statement of the problem, and was therefore not necessarily assignable. Our basic objective is to estimate a linear functional whose coefficient may be chosen without that constraint.

Before attempting to use the class (2.20) to improve the general bounds for linear functionals given in (2.13) and (2.14), we notice one more thing. Addition of the extremum principles to (2.10) and (2.12) eliminates the unknown L_0, but at the expense of giving away the first or second term in the identity

$$0 = (L_\alpha - L_0) + (L_0 - L_\beta) - (L_\alpha - L_\beta) . \tag{2.21}$$

In Fujita's linear theory this is masked by the special orthogonality property (2.19) in the form

$$0 = \langle u_\alpha - u_0, u_\beta - u_0 \rangle = \frac{1}{2}\|u_\alpha - u_0\|^2 + \frac{1}{2}\|u_0 - u_\beta\|^2 - \frac{1}{2}\|u_\alpha - u_\beta\|^2 . \tag{2.22}$$

His weak estimates (1.8) - (1.9) require that the first or second of (2.22) be given away, and do not therefore depend on the orthogonality per se. It is therefore his weak estimates which we shall be trying to generalize when we optimize (2.13) and (2.14).

On the other hand, his strong inequality (1.7) does not give away the stated terms $L_\alpha - L_0$ and $L_0 - L_\beta$, but it does seem to depend critically on the orthogonality property

$$\begin{aligned} \|\tfrac{1}{2}(u_\alpha + u_\beta) - u_0\|^2 - \tfrac{1}{4}\|u_\alpha - u_\beta\|^2 &= \langle u_\alpha - u_0, u_\beta - u_0 \rangle \\ &= (T^*(u_\alpha - u_0), x_0 - x_\beta) = 0 . \end{aligned} \tag{2.23}$$

For this reason we expect his strong inequality to be harder to generalize, even though something may be achieved in particular cases (see § 2(x)).

(vi) Intermediate generality

In seeking to optimize the general bounds (2.13) and (2.14) on linear functionals, we find it illuminating to concentrate first upon some cases which are more general than (1.14) or (1.17), but less general than an arbitrary saddle functional $L[x, u]$. These are separable cases of type

$$L = (x, T^* u) - N(x) + G(u) \tag{2.24}$$

in which $N(x)$ and $G(u)$ are convex functionals of the single variables x and u respectively. In passing we can notice that

$$L = (x, T^* u) - N(x)G(u) \tag{2.25}$$

is concave in x and convex in u, provided both $G \geq 0$ (or N linear) and $N \leq 0$ (or G linear) in addition to the convexity of N and G.

Examples of (2.24) in which one of $N(x)$ or $G(u)$ are linear arise in fields such as network theory and elasticity. In the latter x can be a generalized stress (cf. Sewell 1973a,b) or bending moment entering a convex $N(x)$, with displacement u appearing in a linear $G(u)$.

In the next subsection we carry out the optimization for

$$L = (x, T^* u) + (f, x) + G(u) \tag{2.26}$$

obtained from (1.14) by letting G be any strictly convex functional, instead of quadratic. From (1.13) this generates the problem

$$\begin{aligned} T^* u &= -f\,, \qquad &(\alpha) \\ Tx &= -g(u)\,, \qquad &(\beta) \end{aligned} \tag{2.27}$$

where

$$g(u) = G'(u)\,.$$

A prime will signify gradients of $G(u)$, and also of $g(u)$ below.

The inequality (2.13) reduces to

$$G(u_+) - (x_\beta, T^* u_\beta + f) - G(u_\beta) \geq -(x_0, \frac{\partial L}{\partial x_+}) \tag{2.28}$$

in which

$$\frac{\partial L}{\partial x_+} = T^* u_+ + f, \qquad Tx_\beta = -g(u_\beta) . \tag{2.29}$$

Here u_+ is any element in the domain of T^*.

The inequality (2.14) reduces to

$$G(u_\alpha) - (f, x_-) + \langle u_-, \frac{\partial L}{\partial u_-} - Tx_- \rangle - G(u_-) \geq \langle u_0, \frac{\partial L}{\partial u_-} \rangle \tag{2.30}$$

in which

$$T^* u_\alpha = -f, \qquad \frac{\partial L}{\partial u_-} = Tx_- + g(u_-) . \tag{2.31}$$

Here x_- is any element in the domain of T, and u_- is any element in the domain of $g(u)$.

(vii) Optimization of the first bound

Recalling (2.20), we choose for the u_+ in (2.28) the restricted class

$$u_+ = u_\alpha + kq \tag{2.32}$$

for any q now in the domain F' of T^*, and any scalar k. Then $(2.29)_1$ with $(2.31)_1$ implies

$$\frac{\partial L}{\partial x_+} = kT^* q \tag{2.33}$$

and (2.28) becomes

$$G(u_\alpha + kq) - (x_\beta, T^* u_\beta + f) - G(u_\beta) \geq -k(x_0, T^* q) . \tag{2.34}$$

We now optimize this inequality approximately with respect to k. It will turn out under suitable circumstances that k is small. Acting on this assumption we write

$$G(u_\alpha + kq) = G(u_\alpha) + k\langle g(u_\alpha), q\rangle + \frac{1}{2}k^2\langle g'(u_\alpha)q, q\rangle + 0(k^3) . \tag{2.35}$$

Inserting this in (2.34) and omitting the higher order terms gives

$$\frac{1}{2}C + F_1 k + \frac{1}{2}B_1 k^2 \geq 0 \tag{2.36}$$

with the following shorthand for the coefficients

$$\frac{1}{2}C \equiv G(u_\alpha) - (x_\beta, T^* u_\beta + f) - G(u_\beta) = L_\alpha - L_\beta \geq 0 , \tag{2.37}$$

$$F_1 \equiv \langle g(u_\alpha) - g(u_0), q\rangle = \langle g(u_\alpha), q\rangle + (x_0, T^* q) , \tag{2.38}$$

$$\frac{1}{2}B_1 \equiv \frac{1}{2}\langle q, g'(u_\alpha)q\rangle . \tag{2.39}$$

A sufficient condition for the strict convexity of $G(u)$ can be given in terms of a mean value theorem (see § 3), and implies that

$$C \equiv \langle u_\alpha - u_\beta, g'(\bar{u})(u_\alpha - u_\beta)\rangle > 0 \tag{2.40}$$

when u_α and u_β are distinct, where the operator $g'(\bar{u})$ is evaluated at some intermediate $\bar{u}$ between u_α and u_β. It also implies the strict inequality

$$B_1 > 0 \tag{2.41}$$

for $q \neq 0$.

Under these two strict inequalities, we optimize (2.36) with respect to k by considering two cases.

(a) $k > 0$ implies

$$\frac{C}{k} + 2F_1 + B_1 k \geq 0 . \tag{2.42}$$

The left side is least, e.g. by completing the square, at $k = (C/B_1)^{\frac{1}{2}}$ and the best result is

$$F_1 + (CB_1)^{\frac{1}{2}} \geq 0 \ . \tag{2.43}$$

(b) $k < 0$ implies

$$0 \geq \frac{C}{k} + 2F_1 + B_1 k \ . \tag{2.44}$$

The right side is greatest at $k = -(C/B_1)^{\frac{1}{2}}$, and the best result is

$$0 \geq F_1 - (CB_1)^{\frac{1}{2}} \ . \tag{2.45}$$

Inequalities (2.43) and (2.45) for the objective linear functional

$$(x_0, T^* q) = -\langle g(u_0), q \rangle \tag{2.46}$$

of the solution variable x_0 can be summarized as

$$|\langle g(u_\alpha), q\rangle + (x_0, T^* q)| \leq (CB_1)^{\frac{1}{2}} \ . \tag{2.47}$$

It has to be remembered that this result is not rigorous because higher order terms were omitted in going from (2.35) to (2.36). Rigorous bounds can be obtained by inserting

$$k = \pm (C/B_1)^{\frac{1}{2}} \tag{2.48}$$

into (2.34). These values of k will not in general provide the best bounds on (2.46), but if $u_\alpha \simeq u_\beta$ they will be close to the optimum bounds because $C = 2(L_\alpha - L_\beta)$ will then be small. Without loss of generality we can assume that B_1 is of order unity so that the resulting k is

small, and the neglect of higher order terms in (2.36) will be justified. Obviously the best bounds are obtained when u_α and u_β are as nearly equal as possible, since as $u_\alpha \to u_\beta$, $C \to 0$ and the left side of (2.47) tends to zero.

The linear problem (1.2) is recovered with

$$G(u) = \frac{1}{2}\langle u,u\rangle, \quad g(u) = u, \quad g'(u) = I \tag{2.49}$$

so that

$$C = \|u_\alpha - u_\beta\|^2, \quad B_1 = \|q\|^2 .$$

The result (2.47) is then exact, namely

$$|\langle u_\alpha - u_0, q\rangle| \le \|u_\alpha - u_\beta\| \, \|q\| \tag{1.9}$$

which is one of Fujita's weak estimates.

(viii) Optimization of the second bound

Inequality (2.30) is the basis for the second bound, and it involves both x_- and u_-. Again recalling (2.20) we choose the class of points

$$x_- = x_\beta + hp, \qquad u_- = u_\beta + kq \tag{2.50}$$

for any p in the domain E' of T, any q such that u_- (like u_β) is in the domain of $g(u)$, and any scalars h and k.

Insertion into $(2.31)_2$, expanding and using $(2.29)_2$ implies

$$\frac{\partial L}{\partial u_-} = hTp + kg'(u_\beta)q + 0(k^2) \tag{2.51}$$

with an obvious extension of the $0(k^2)$ notation. Here $g'(u_\beta)$ is an

operator acting on q. The inequality (2.30) expanded about the β-point x_β, u_β becomes, after omission of $0(k^3)$ terms,

$$\frac{1}{2}C + F_2 k + \frac{1}{2}B_2 k^2 \geq 0 \tag{2.52}$$

because the terms in h cancel exactly. Here $C = 2(L_\alpha - L_\beta)$ as in (2.37), but the other coefficients are now

$$F_2 \equiv \langle u_\beta - u_0, g'(u_\beta)q \rangle , \tag{2.53}$$

$$\frac{1}{2}B_2 \equiv \frac{1}{2}\langle q, g'(u_\beta)q \rangle . \tag{2.54}$$

The form of these shows that it would be enough to regard $g'(u_\beta)q = Q$ (say) as arbitrary, i.e. to have begun with $q = [g'(u_\beta)]^{-1}Q$ in (2.50).

Optimizing (2.52) exactly as before, we find

$$|\langle u_\beta - u_0, Q \rangle| \leq (CB_2)^{\frac{1}{2}} \tag{2.55}$$

in place of (2.47). Again this is approximate, but rigorous bounds can be obtained by using

$$k = \pm (C/B_2)^{\frac{1}{2}} \tag{2.56}$$

to deduce exact values from (2.50) for substitution in (2.30).

In the linear problem (1.2) we have only to put $g'(u) = I$ in (2.53) and (2.54) to see that (2.55) becomes

$$|\langle u_\beta - u_0, q \rangle| \leq \|u_\alpha - u_\beta\| \, \|q\| \tag{1.8}$$

which is Fujita's other weak estimate.

(iv) Optimization in the general case for weak bounds

Here we indicate the optimization procedure that would be required in the general case for weak bounds, i.e. as it would apply to (2.13) and (2.14).

At certain points we shall need to suppose that a 'Taylor' expansion of the following type can be employed (see also the discussion in § 3(i)):

$$L[x+\xi,\ u+\upsilon] = L[x,u] + (\xi,\ \frac{\partial L}{\partial x}) + \langle \upsilon,\ \frac{\partial L}{\partial u}\rangle$$

$$+ \frac{1}{2}(\xi,\ \frac{\partial^2 L}{\partial x^2}\xi) + (\xi,\ \frac{\partial^2 L}{\partial x \partial u}\upsilon) + \frac{1}{2}\langle \upsilon,\ \frac{\partial^2 L}{\partial u^2}\upsilon\rangle$$

$$+ \text{higher order terms.} \qquad (2.57)$$

In this expansion on the product space $E \times F$, the linear operators act on the elements of E or F which follow them as before, and with the assumed property that

$$(\xi,\ \frac{\partial^2 L}{\partial x \partial u}\upsilon) = \langle \upsilon,\ \frac{\partial^2 L}{\partial u \partial x}\xi\rangle . \qquad (2.58)$$

This last property is exemplified by the adjointness statement (1.4) for the particular bilinear functional $L = (x, T^* u) = \langle u, Tx\rangle$. In general $\partial^2 L/\partial x \partial u \neq \partial^2 L/\partial u \partial x$, as $T \neq T^*$ illustrates. It is assumed that $\partial^2 L/\partial x^2$ and $\partial^2 L/\partial u^2$ are self-adjoint.

First we wish to optimize (2.13) with respect to all plus points in a suitably chosen family. We consider a family centered on the α-point, i.e. we choose (cf. (2.20) and (2.32))

$$x_+ = x_\alpha + hp, \qquad u_+ = u_\alpha + kq \tag{2.59}$$

and optimize with respect to the scalars h and k.

The expansion of L_+ according to (2.57), and the operator expansion of $\partial L/\partial x_+$, are simplified because $\partial L/\partial x_\alpha = 0$ by (2.2). They are

$$L_+ = L_\alpha + k\langle q, \frac{\partial L}{\partial u_\alpha}\rangle + \frac{1}{2}h^2(p, \frac{\partial^2 L}{\partial x_\alpha^2}p) + hk(p, \frac{\partial^2 L}{\partial x \partial u_\alpha}q) + \frac{1}{2}k^2\langle q, \frac{\partial^2 L}{\partial u_\alpha^2}q\rangle + \cdots \tag{2.60}$$

and

$$\frac{\partial L}{\partial x_+} = h\frac{\partial^2 L}{\partial x_\alpha^2}p + k\frac{\partial^2 L}{\partial x \partial u_\alpha}q + \cdots . \tag{2.61}$$

The dots will consistently denote higher order terms in h and k. Using (2.61) to eliminate the cross-derivatives from (2.60), and substituting the result back into (2.13) gives

$$L_\alpha - L_\beta - (x_\alpha - x_0, \frac{\partial L}{\partial x_+}) + k\langle q, \frac{\partial L}{\partial u_\alpha}\rangle - \frac{1}{2}h^2\left(p, \frac{\partial^2 L}{\partial x_\alpha^2}p\right) + \frac{1}{2}k^2\langle q, \frac{\partial^2 L}{\partial u_\alpha^2}q\rangle + \cdots \geq 0 . \tag{2.62}$$

We can also expand $\partial L/\partial u_\alpha$ about the solution point x_0, u_0, giving

$$\frac{\partial L}{\partial u_\alpha} = \frac{\partial^2 L}{\partial u \partial x_0}(x_\alpha - x_0) + \frac{\partial^2 L}{\partial u_0^2}(u_\alpha - u_0) + \cdots \tag{2.63}$$

since $\partial L/\partial u_0 = 0$. Insert this into (2.62), together with (2.61). If we assume that all second order derivatives can be taken at x_α, u_α instead

of x_0, u_0, to the order of accuracy retained,

$$L_\alpha - L_\beta - h\langle x_\alpha - x_0, \frac{\partial^2 L}{\partial x_\alpha^2} p\rangle + k\langle u_\alpha - u_0, \frac{\partial^2 L}{\partial u_\alpha^2} q\rangle$$

$$- \frac{1}{2} h^2\langle p, \frac{\partial^2 L}{\partial x_\alpha^2} p\rangle + \frac{1}{2} k^2\langle q, \frac{\partial^2 L}{\partial u_\alpha^2} q\rangle + \cdots \geq 0 . \tag{2.64}$$

Omitting now the higher order terms, (2.64) is of the form

$$C + 2F_1 k + B_1 k^2 + 2G_1 h + A_1 h^2 \geq 0 \tag{2.65}$$

in which $\frac{1}{2} C = L_\alpha - L_\beta \geq 0$ by the dual extremum principles, as before. Also $B_1 \geq 0$ and $A_1 \geq 0$ because L is concave in x and convex in u.

If $\partial^2 L/\partial x_\alpha^2 = 0$ so that $A_1 = G_1 = 0$, or if we arbitrarily set $h = 0$, we can choose k in (2.65) to find optimum bounds for F_1 as in (2.36), and this gives

$$\left|\langle u_\alpha - u_0, \frac{\partial^2 L}{\partial u_\alpha^2} q\rangle\right| \leq \left[2(L_\alpha - L_\beta)\langle q, \frac{\partial^2 L}{\partial u_\alpha^2} q\rangle\right]^{\frac{1}{2}} . \tag{2.66}$$

Similarly, if $\partial^2 L/\partial u_\alpha^2$ so that $B_1 = F_1 = 0$, or if we arbitrarily set $k = 0$, we can choose h to obtain optimum bounds for G_1 as

$$\left|\langle x_\alpha - x_0, \frac{\partial^2 L}{\partial x_\alpha^2} p\rangle\right| \leq \left[-2(L_\alpha - L_\beta)\langle p, \frac{\partial^2 L}{\partial x_\alpha^2} p\rangle\right]^{\frac{1}{2}} . \tag{2.67}$$

If $A_1 > 0$ and $B_1 > 0$ the inequality (2.65) can be rearranged in the form

$$A_1\left(h - \frac{G_1}{A_1}\right)^2 + B_1\left(k - \frac{F_1}{B_1}\right)^2 + C - \frac{G_1^2}{A_1} - \frac{F_1^2}{B_1} \geq 0 \tag{2.68}$$

and the choice $h = G_1/A_1$, $k = F_1/B_1$ leads to

$$A_1B_1C \geq A_1F_1^2 + B_1G_1^2 \ . \tag{2.69}$$

Inequalities (2.66) and (2.67) can be deduced from (2.69).

A similar general analysis can be carried out to optimize (2.14) by expanding x_-, u_- about the β-point, in place of (2.59).

(x) Strong bounds

We return to the remark made after (2.23). When it turns out that L_0 happens to be a linear functional of x_0 or u_0, it may be possible to build estimates on (2.10) or (2.12), without needing to give away the additional quantities $L_0 - L_\beta$ or $L_\alpha - L_0$ required to arrive at (2.13) and (2.14). Such estimates can be called 'strong' in the sense of (1.7). We may therefore envisage the optimization of (2.10) and (2.12) in such special uses, using ideas associated with those described for (2.13) and (2.14).

In the linear problem (1.2), we can see from (1.6) that

$$L_0 = \frac{1}{2}(f, x_0) \tag{2.70}$$

and (2.10) reduces to

$$\frac{1}{2}\langle u_+, u_+\rangle \geq -(x_0, T^*u_+ + \frac{1}{2}f) \tag{2.71}$$

for any u_+ in the domain of T^*. The appropriate optimization no longer involves an expansion about either the α-point (as in (2.32)) or the β-point (as in (2.50)), but about another point in the family (2.20) midway between them, i.e.

$$u_+ = \frac{1}{2}(u_\alpha + u_\beta) + kq \ . \tag{2.73}$$

Fujita's strong estimate (1.7) can now be recovered by optimizing (2.71) with respect to the k of (2.73).

Nothing new is achieved from (2.12), since this reduces to

$$\frac{1}{2}\langle u_-, u_-\rangle \geq -(x_0, T^* u_- + \frac{1}{2} f) \tag{2.74}$$

for any u_- in the domain of T^*, which is therefore exactly the same as (2.71).

Instead of offering more general theory, we give the explicit example described in §§6(ii) - (v), in which L_0 has the linear form (6.22).

3. Basic Theoretical Framework

(i) Mean value theorems

Underlying all of our general theory is a result whose essence was stated in equation (5.3) of our earlier joint paper (op. cit.). To begin with let $L[x,u]$ be a function of two real variables, defined over a rectangular domain which allows us to join an arbitrary pair of points x_+, u_+ and x_-, u_- by a two-segment path parallel to the axes and lying entirely within the domain, as in Fig. 3.1.

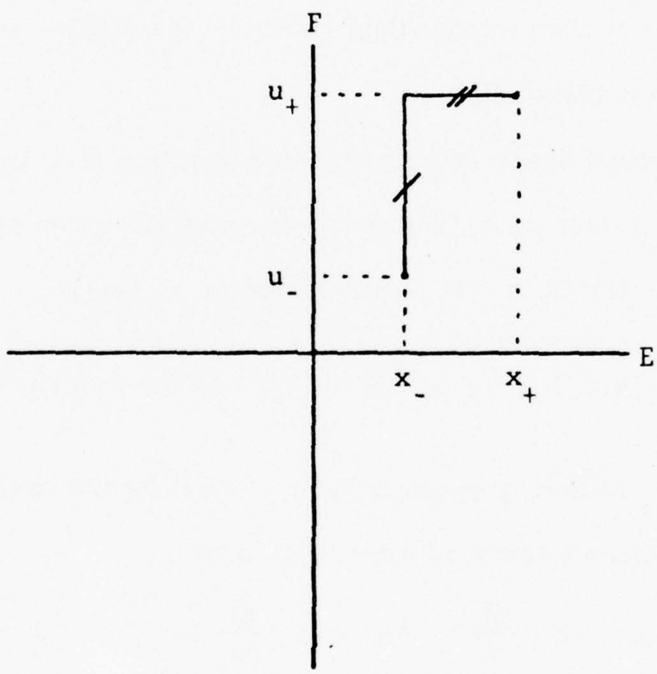


Fig. 3.1. Mean value theorem route

Then two uses of the second mean value theorem for one variable, namely u on the single-barred segment and x on the double-barred segment, give

$$L_+ - L_- - (x_+ - x_-)\frac{\partial L}{\partial x_+} - (u_+ - u_-)\frac{\partial L}{\partial u_-}$$

$$= \frac{1}{2}(u_+ - u_-)^2 \overline{\frac{\partial^2 L}{\partial u^2}} - \frac{1}{2}(x_- - x_+)^2 \overline{\overline{\frac{\partial^2 L}{\partial x^2}}} . \tag{3.1}$$

The bar and double-bar over the second derivatives denote evaluation at different and unknown intermediate points on the single- and double-barred segments respectively.

In our function space setting we can generalize (3.1) by using the abstract form of Taylor series with integral remainder given by, for instance, Cartan (1971), p. 70, and Rall (1969), p. 124:

$$f(a + h) = f(a) + f'(a) \cdot h + \int_0^1 (1 - t)f''(a + th)(h)^2 dt \ .$$

This leads to the following generalization of (3.1) for the case of the two inner product spaces introduced after (1.2) to be

$$L_+ - L_- - (x_+ - x_-, \frac{\partial L}{\partial x_+}) - \langle u_+ - u_-, \frac{\partial L}{\partial u_-} \rangle$$

$$= \frac{1}{2}\langle u_+ - u_-, \overline{\frac{\partial^2 L}{\partial u^2}}(u_+ - u_-)\rangle - \frac{1}{2}(x_+ - x_-, \overline{\overline{\frac{\partial^2 L}{\partial x^2}}}(x_+ - x_-)) , \tag{3.2}$$

where

$$\overline{\frac{\partial^2 L}{\partial u^2}} = \int_0^1 (1 - t) \frac{\partial^2 L}{\partial u^2} (x_-, u_- + t(u_+ - u_-))dt$$

$$\overline{\overline{\frac{\partial^2 L}{\partial x^2}}} = \int_0^1 (1 - t) \frac{\partial^2 L}{\partial x^2} (x_+ + t(x_- - x_+), u_+)dt .$$

In specific applications these can be replaced by the appropriate mean-value theorem. The expression on the left in (3.2) corresponds to the saddle quantity appearing in (2.1), but is not now assumed necessarily to be one-signed. The symbols reminiscent of second derivatives on the right have now to be regarded as linear operators acting on the elements of F or E which follow them, as for $g'(u)$ in §§ 2(vii) and (viii). Further details on higher derivatives in vector spaces are described by Rall (1969, §§ 18, 19).

Note that a bilinear functional, like $(x, T^* u)$ or the Lagrangian which generates linear programming (Noble and Sewell, op. cit., § 10(ii)), has only mixed second derivatives, and so contributes identically zero to the saddle quantity. This is evident from the right side of (3.2), and can be verified on the left side by direct substitution.

For certain purposes, ultimately connected with embedded problems, we shall also draw conclusions from mean value statements for gradients of the type

$$\frac{\partial L}{\partial x_+} - \frac{\partial L}{\partial x_-} = \overline{\overline{\frac{\partial^2 L}{\partial x^2}}}(x_+ - x_-) + \overline{\frac{\partial^2 L}{\partial x \partial u}}(u_+ - u_-)\,, \tag{3.3}$$

$$\frac{\partial L}{\partial u_+} - \frac{\partial L}{\partial u_-} = \overline{\overline{\frac{\partial^2 L}{\partial u \partial x}}}(x_+ - x_-) + \overline{\frac{\partial^2 L}{\partial u^2}}(u_+ - u_-)\,. \tag{3.4}$$

Precise statements about mean value theorems for operators are given by Rall (op. cit., § 20). Examples of second derivative operators which happen also to be constant are found in the problems generated by (1.14) and (1.17), where

$$\begin{aligned} \frac{\partial^2 L}{\partial x^2} &= -\rho I, & \frac{\partial^2 L}{\partial u^2} &= I\,, \\ \frac{\partial^2 L}{\partial x \partial u} &= T^*, & \frac{\partial^2 L}{\partial u \partial x} &= T\,. \end{aligned} \tag{3.5}$$

<u>(ii) Boundedness hypotheses</u>

We have now removed the saddle hypothesis (2.1), and in its place we begin to build our theoretical framework upon the following boundedness hypotheses. We suppose that there exists a rectangular domain of the product space $E \times F$ in which real numbers k_x, k_u, or K_x, K_u, or both pairs, can be found such that

(a) for each given u_+ and every pair x_+, x_-

$$K_x \|x_+ - x_-\|^2 \geq -\left(x_+ - x_-, \overline{\overline{\frac{\partial^2 L}{\partial x^2}}}(x_+ - x_-)\right) \geq k_x \|x_+ - x_-\|^2\,, \tag{3.6}$$

(b) for each given x_- and every pair u_+, u_-

$$K_u \|u_+ - u_-\|^2 \geq \left\langle u_+ - u_-, \overline{\frac{\partial^2 L}{\partial u^2}}(u_+ - u_-)\right\rangle \geq k_u \|u_+ - u_-\|^2\,. \tag{3.7}$$

In general it may be that k_x and K_x depend on u_+, and that k_u and K_u depend on x_-, and we sometimes emphasize this by writing

$$k_x(u_+),\ K_x(u_+),\ k_u(x_-),\ K_u(x_-)\ . \tag{3.8}$$

Of course they may sometimes be constant over the domain. We allow that these bounds (3.8) may have either sign. For quadratic functionals such as (1.17) there exist the trivial bounds

$$K_x = k_x = \rho, \qquad K_u = k_u = 1\ . \tag{3.9}$$

Evidently (3.2) with (3.6) and (3.7) may be written

$$B_{+-} \geq L_+ - L_- - (x_+ - x_-, \frac{\partial L}{\partial x_+}) - \langle u_+ - u_-, \frac{\partial L}{\partial u_-} \rangle \geq b_{+-} \tag{3.10}$$

with the shorthand

$$\begin{aligned} B_{+-} &\equiv \frac{1}{2} K_x(u_+) \|x_+ - x_-\|^2 + \frac{1}{2} K_u(x_-) \|u_+ - u_-\|^2\ , \\ b_{+-} &\equiv \frac{1}{2} k_x(u_+) \|x_+ - x_-\|^2 + \frac{1}{2} k_u(x_-) \|u_+ - u_-\|^2\ . \end{aligned} \tag{3.11}$$

Sufficient conditions for $L[x,u]$ to be a saddle functional concave in x and convex in u are that

$$k_x(u_+) \geq 0 \quad \text{and} \quad k_u(x_-) \geq 0 \tag{3.12}$$

over a rectangular domain, so that $b_{+-} \geq 0$. Then (2.1) follows from (3.10) without need of K_x and K_u. [Alternatively L is convex in x and concave in u if $0 \geq K_x$ and $0 \geq K_u$, without need of k_x and k_u. But this can be reduced to the first case by turning the saddle functional upside down]. Sufficient conditions for a strict saddle are strict inequalities

in (3.12), and then any solution x_0, u_0 of (1.13) is unique (Sewell, 1969, inequality (2.52)).

The type of theory presented in § 2 is available when (3.12) hold with at least one of the inequalities strict.

However, there are many situations when such sufficient hypotheses fail, such as nonlinear elasticity or nonconvex optimization, and in which there may still be a need to bound linear functionals. There will now be a different linear functional to bound for each different solution, and the needed alternative hypotheses will reflect the attention which must be given to domain boundaries separating the individual solutions.

Consider the example of a function of two real variables

$$L = -(\tfrac{1}{4}x^4 + \tfrac{1}{2}ax^2 + bx + c)u - px \tag{3.13}$$

with scalar coefficients a, b, c, p. For this $K_u = k_u = 0$ and (3.10) reduces to

$$K_x \geq (3\bar{x}^2 + a)u_+ \geq k_x \tag{3.14}$$

for some mean $\bar{x}$ between x_+ and x_-. Evidently in the half-space $u_+ > 0$

$$k_x(u_+) = au_+, \qquad K_x \text{ does not exist}, \tag{3.15}$$

whereas in $u_+ < 0$ k_x does not exist but $K_x = au_+$. Therefore a sufficient condition for (3.13) to be a saddle function strictly concave in x

$$\text{in } u_+ > 0 \text{ is } a > 0. \tag{3.16}$$

This illustrates (2.25).

When $a < 0$, however, the equation $\partial L/\partial u = 0$ is a quartic in x which might not have a unique solution. Its solutions will be separated by the turning points of the quartic, which we specified by the roots of

$$\frac{\partial^2 L}{\partial u \partial x} = 0 \quad \text{i.e.} \quad x^3 + ax + b = 0 \ . \tag{3.17}$$

The required new hypotheses would avoid such roots. It is no accident that (3.17) is also the equilibrium surface for the cusp catastrophe (e.g. see Sewell, 1976, for diagrams and mechanical examples), and further insight can be obtained by pursuing this connection.

At this point, however, we have said enough to motivate the following choice of alternative boundedness hypotheses required when the sufficient saddle hypotheses (3.12) fail. We assume that in a rectangular domain there exist real numbers

$$c_x(u_+) \geq 0 \quad \text{and/or} \quad c_u(x_-) \geq 0 \tag{3.18}$$

depending possibly on u_+ and x_- as indicated, such that for every pair of points

$$\left\| \overline{\overline{\frac{\partial^2 L}{\partial u \partial x}}} (x_+ - x_-) \right\| \geq c_x \|x_+ - x_-\| \ , \tag{3.19}$$

$$\left\| \overline{\frac{\partial^2 L}{\partial x \partial u}} (u_+ - u_-) \right\| \geq c_u \|u_+ - u_-\| \ . \tag{3.20}$$

If it is also the case that there exist real numbers

$$d_x(u_+) \geq 0 \quad \text{and/or} \quad d_u(x_-) \geq 0 \ , \tag{3.21}$$

again depending possibly on u_+ and x_-, such that for every pair of points

$$d_x \|x_+ - x_-\| \geq \left\| \overline{\frac{\partial^2 L}{\partial x^2}} (x_+ - x_-) \right\| , \tag{3.22}$$

$$d_u \|u_+ - u_-\| \geq \left\| \overline{\frac{\partial^2 L}{\partial u^2}} (u_+ - u_-) \right\| , \tag{3.23}$$

then Schwarz's inquality allows the values

$$-k_x = K_x = d_x \geq 0 \tag{3.24}$$

and

$$-k_u = K_u = d_u \geq 0 \tag{3.25}$$

for the numbers (3.8). Therefore (3.22) and (3.23) can be used when the sufficient saddle requirement (3.12) fails.

<u>(iii) Error estimates</u>

We draw some conclusions from (3.10), first of ail without any assumption about the signs of k_x, k_u, K_x, K_u.

(a) Choosing $x_+, u_+ = x_\alpha, u_\alpha$ and $x_-, u_- = x_0, u_0$ (as in (2.3) but without the saddle hypothesis) implies

$$B_{\alpha 0} \geq L_\alpha - L_0 \geq b_{\alpha 0} . \tag{3.26}$$

(b) Choosing $x_+, u_+ = x_0, u_0$ and $x_-, u_- = x_\beta, u_\beta$ (as in (2.4) but without the saddle hypothesis) implies

$$B_{0\beta} \geq L_0 - L_\beta \geq b_{0\beta} . \tag{3.27}$$

(c) Choosing $x_+, u_+ = x_\alpha, u_\alpha$ and $x_-, u_- = x_\beta, u_\beta$ implies

$$B_{\alpha\beta} \geq L_\alpha - L_\beta \geq b_{\alpha\beta} . \tag{3.28}$$

Then $(3.26)_1 + (3.27)_1$ with $(3.28)_2$ implies

$$B_{\alpha 0} + B_{0\beta} \geq L_\alpha - L_\beta \geq b_{\alpha\beta} \tag{3.29}$$

and $(3.26)_2 + (3.27)_2$ with $(3.28)_1$ implies

$$B_{\alpha\beta} \geq L_\alpha - L_\beta \geq b_{\alpha 0} + b_{0\beta} . \tag{3.30}$$

These last two inequalities, with (3.11), can be regarded as composite error estimates for the solution quantities in the left of (3.29) or the right of (3.30). For example, the latter written explicitly is

$$\begin{aligned} L_\alpha - L_\beta \geq \frac{1}{2} k_x(u_\alpha) \|x_\alpha - x_0\|^2 + \frac{1}{2} k_x(u_0) \|x_\beta - x_0\|^2 \\ + \frac{1}{2} k_u(x_0) \|u_\alpha - u_0\|^2 + \frac{1}{2} k_u(x_\beta) \|u_\beta - u_0\|^2 . \end{aligned} \tag{3.31}$$

In the case of a saddle functional satisfying (3.12) with

$$k_x(u_\alpha) > 0 \quad \text{and/or} \quad k_u(x_\beta) > 0 \tag{3.32}$$

more can be given away from (3.31) to imply

$$\frac{2}{k_x(u_\alpha)} (L_\alpha - L_\beta) \geq \|x_\alpha - x_0\|^2 \tag{3.33}$$

and/or

$$\frac{2}{k_u(x_\beta)} (L_\alpha - L_\beta) \geq \|u_\beta - u_0\|^2 . \tag{3.34}$$

These simple error estimates are the appropriate generalizations of (2.8) and they improve results given by Zago (1976, Chapter 2).

Next suppose that, instead of the sufficient requirement (3.12) for a saddle functional, both (3.19) - (3.20) and (3.22) - (3.23) hold. Then the triangle inequality applied to the mean value statements (3.3) and (3.4) leads to

$$-d_x(u_+)\,\|x_+ - x_-\| + c_u(x_-)\,\|u_+ - u_-\| \le \left\|\frac{\partial L}{\partial x_+} - \frac{\partial L}{\partial x_-}\right\| , \tag{3.35}$$

$$c_x(u_+)\,\|x_+ - x_-\| - d_u(x_-)\,\|u_+ - u_-\| \le \left\|\frac{\partial L}{\partial u_+} - \frac{\partial L}{\partial u_-}\right\| . \tag{3.36}$$

Because the basic problem (1.13) is stated in terms of gradients (satisfying also (2.2)), the right sides of (3.35) and (3.36) can be regarded as *known*, *in particular* under choices of the disposable plus and minus points like those in (a) - (c) above. Therefore these inequalities are the basis for another class of error estimates different from (3.33) and (3.34). Their usefulness may depend somewhat on the extent to which c_x, c_u, d_x, d_u are actually constant over the considered rectangular domain.

In any event, if it is also true that

$$c_x c_u - d_x d_u > 0 , \tag{3.37}$$

(3.35) and (3.36) can be solved to give

$$\|x_+ - x_-\| \le \frac{1}{c_x c_u - d_x d_u}\left[d_u \left\|\frac{\partial L}{\partial x_+} - \frac{\partial L}{\partial x_-}\right\| + c_u \left\|\frac{\partial L}{\partial u_+} - \frac{\partial L}{\partial u_-}\right\|\right], \tag{3.38}$$

$$\|u_+ - u_-\| \le \frac{1}{c_x c_u - d_x d_u}\left[c_x \left\|\frac{\partial L}{\partial x_+} - \frac{\partial L}{\partial x_-}\right\| + d_x \left\|\frac{\partial L}{\partial u_+} - \frac{\partial L}{\partial u_-}\right\|\right]. \tag{3.39}$$

These can be substituted back, with (3.24) and (3.25), into (3.20) to give an upper bound for B_{+-} and a lower bound for b_{+-}. They can also be substituted, after Schwarz's inequality, into either or both of the inner products appearing in the centre expression of (3.10), finally giving bounds for what is left there.

For example, we can add

$$\|x_+ - x_-\| \left\|\frac{\partial L}{\partial x_+}\right\| \ge (x_+ - x_-, \frac{\partial L}{\partial x_+}) \ge - \|x_+ - x_-\| \left\|\frac{\partial L}{\partial x_+}\right\| \tag{3.40}$$

to (3.10), and then substitute (3.38) and (3.39) into both of the resulting bounds, giving

$$\begin{aligned} &|L_+ - L_- - \langle u_+ - u_-, \frac{\partial L}{\partial u_-}\rangle| \\ &\le \frac{\left\|\frac{\partial L}{\partial x_+}\right\|}{c_x c_u - d_x d_u}\left[d_u \left\|\frac{\partial L}{\partial x_+} - \frac{\partial L}{\partial x_-}\right\| + c_u \left\|\frac{\partial L}{\partial u_+} - \frac{\partial L}{\partial u_-}\right\|\right] \\ &+ \frac{1}{2}\frac{d_x}{(c_x c_u - d_x d_u)^2}\left[d_u \left\|\frac{\partial L}{\partial x_+} - \frac{\partial L}{\partial x_-}\right\| + c_u \left\|\frac{\partial L}{\partial u_+} - \frac{\partial L}{\partial u_-}\right\|\right]^2 \\ &+ \frac{1}{2}\frac{d_u}{(c_x c_u - d_x d_u)^2}\left[c_x \left\|\frac{\partial L}{\partial x_+} - \frac{\partial L}{\partial x_-}\right\| + d_x \left\|\frac{\partial L}{\partial u_+} - \frac{\partial L}{\partial u_-}\right\|\right]^2 . \end{aligned} \tag{3.41}$$

In the next Section we give an example of this result.

4. General Embedding Method

(i) Embedding

We consider now a problem given ab initio in the form (1.19). We can allow that it be decomposable or nondecomposable, linear or nonlinear. Identify the variable in the problem with the x of an inner product space E, and suppose the operator N ranges in a second inner product space F. The typical element u of F is employed in the role of a Lagrangian multiplier by constructing the functional

$$L[x,u] = -\langle u, N(x)\rangle + (p,x) \tag{4.1}$$

where p is an assignable vector in E.

Suppose the adjoint N'^{*} of the Gateaux differential N' exists (Barnsley and Robinson (1976) give technical details in the case of two Hilbert spaces). Then the gradients of L are

$$\frac{\partial L}{\partial x} = -N'^{*}(x)u + p, \quad \frac{\partial L}{\partial u} = -N(x) . \tag{4.2}$$

Then (1.13) in the form

$$N'^{*}(x)u - p = 0 \qquad (\alpha), \qquad N(x) = 0 \qquad (\beta) \tag{4.3}$$

contains (1.19) embedded as (4.3β), with (4.3α) as an auxiliary equation.

The real objective now is to estimate the linear functional (p, x_β), since x_β is a solution of the ab initio problem, and p is an assignable vector.

The significance of the result (3.41) for this purpose is that because u appears linearly in (4.1) as a Lagrangian multiplier, the quantity

estimated on the left of (3.41) is

$$L_+ - L_- - \langle u_+ - u_-, \frac{\partial L}{\partial u_-} \rangle = -\langle u_+, N(x_+) - N(x_-)\rangle + (p, x_+ - x_-) . \tag{4.4}$$

Therefore the choice

$$x_+, u_+ \quad \text{arbitrary}, \quad x_- = x_\beta \tag{4.5}$$

introduces the objective functional (p, x_β) directly into (4.4), which becomes

$$-\langle u_+, N(x_+)\rangle + (p, x_+) - (p, x_\beta) \tag{4.6}$$

since $N(x_\beta) = 0$. The first two terms in (4.6) are assignable, so (3.41) gives an estimate for (p, x_β) provided the hypotheses leading to (3.41) can be verified.

The linearity of (4.1) in u allows

$$d_u = 0 \tag{4.7}$$

in (3.23) and (3.25). With

$$d_x = \|u_+\| d > 0 \tag{4.8}$$

in (3.24), the constant d corresponds via (3.22) to a bound imposed by Barnsley and Robinson (op. cit.) on the second derivative of the operator $N(x)$.

Suppose there exists

$$c_x(u_+) > 0 \tag{4.9}$$

such that, for all x_+, u_+

$$\|N(x_+)\| = \|N(x_+) - N(x_\beta)\| \geq c_x \|x_+ - x_\beta\| \tag{4.10}$$

so that c_x in (3.19) is effectively a bound on the first derivative of $N(x)$.

Because of (4.7) neither c_u nor (3.20) are required, and with (4.5) the right side of (3.41) reduces to

$$\frac{1}{c_x}\left\|\frac{\partial L}{\partial x_+}\right\|\left\|\frac{\partial L}{\partial u_+}\right\| + \frac{d}{2c_x^2}\|u_+\|\left\|\frac{\partial L}{\partial u_+}\right\|^2 . \tag{4.11}$$

Substituting (4.2) and recalling (4.6) leads finally to the explicit estimate

$$\begin{aligned} &|-\langle u_+, N(x_+)\rangle + (p,x_+) - (p,x_\beta)| \\ &\le \frac{1}{c_x}\|N'^*(x_+)u_+ - p\|\,\|N(x_+)\| + \frac{d}{2c_x^2}\|u_+\|\,\|N(x_+)\|^2 . \end{aligned} \tag{4.12}$$

This corresponds to the result of Barnsley and Robinson (op. cit., inequalities (3.6)), but is obtained here from a different viewpoint.

<u>(ii) Example</u>

We can indicate very briefly how the procedure works by referring to the algebraic example (3.13). Then (4.3) becomes

$$(x^3 + ax + b)u = p \quad (\alpha), \qquad \frac{1}{4}x^4 + \frac{1}{2}ax^2 + bx + c = 0 \quad (\beta) . \tag{4.13}$$

To satisfy (4.10) we have to distinguish not more than four domains of the x-axis, separated by the stationary points of the quartic. In any such fixed open domain the slope (and therefore the cubic coefficient of u in (4.13α)) is nonzero, and a bound c_x for it can be determined. Recalling (3.14), the right side of (3.22) is

$$|(3\bar{x}^2 + a)u_+(x_+ - x_-)| \le |x_+ - x_-| \cdot |u_+| \cdot \max|3x^2 + a| , \tag{4.14}$$

and therefore in (4.8) we choose $d = \max|3x^2 + a|$ over the domain. If

there is a solution x_β of the quartic in that domain, a bound for px_β can be obtained from (4.12) with any u_+ and any x_+. The first term on the right of (4.12) can be made to vanish if we choose the arbitrary

$$x_+, u_+ = x_\alpha, u_\alpha, \quad \text{i.e.} \quad (x_\alpha^3 + ax_\alpha + b)u_\alpha = p \tag{4.15}$$

but this is not essential. Improvement of the bounds is another matter, however, and Barnsley and Robinson (op. cit.) mention the connection with Newton's method. They discuss a particular case of this example in which $a = 0$, $b = c = -\frac{1}{4}$, for which the quartic is actually convex.

5. Nonlinear Programming Method

(i) Governing conditions

Suppose that the basic problem (1.13) is now replaced by a new problem governed by the following different conditions, but again generated from a given scalar functional $L[x,u]$ of the elements of two-inner product spaces E and F.

$$\left.\begin{aligned} \frac{\partial L}{\partial x} &\leq 0 , && (\alpha) \\ x &\geq 0 , && (\beta) \\ \left(x, \frac{\partial L}{\partial x}\right) &= 0 , && \end{aligned}\right\} \qquad (5.1)$$

$$\left.\begin{aligned} \frac{\partial L}{\partial u} &\geq 0 , && (\beta) \\ u &\geq 0 , && (\alpha) \\ \left\langle u, \frac{\partial L}{\partial u}\right\rangle &= 0 . && \end{aligned}\right\} \qquad (5.2)$$

The presence of inequalities of course implies that the elements of E and F are built up ultimately from real numbers (e.g. via the individual entries in real matrices), to which the inequalities are applied. In other words, all elements are ordered so that the inequalities are defined. These governing conditions have again been divided into two subsets labelled (α) and (β) (and a third unlabelled subset, of 'orthogonality conditions'). A point x_α, u_α now denotes any solution of $(5.1\alpha) + (5.2\alpha)$, and a point x_β, u_β is any solution of $(5.1\beta) + (5.2\beta)$.

(ii) Dual extremum principles

When $L[x,u]$ is a saddle functional in the sense of (2.1), the choice (2.3) implies the minimum principle

$$L_\alpha - (x_\alpha, \frac{\partial L}{\partial x_\alpha}) - L_0 \geq -(x_0, \frac{\partial L}{\partial x_\alpha}) + \langle u_\alpha, \frac{\partial L}{\partial u_0}\rangle \geq 0 . \tag{5.3}$$

On the other hand, the choice (2.4) in (2.1) implies the maximum principle

$$L_0 - L_\beta + \langle u_\beta, \frac{\partial L}{\partial u_\beta}\rangle \geq -(x_\beta, \frac{\partial L}{\partial x_0}) + \langle u_0, \frac{\partial L}{\partial u_\beta}\rangle \geq 0 . \tag{5.4}$$

Therefore, in place of (2.5) we have the following dual extremum principles

$$L_\alpha - (x_\alpha, \frac{\partial L}{\partial x_\alpha}) \geq L_0 \geq L_\beta - \langle u_\beta, \frac{\partial L}{\partial u_\beta}\rangle \tag{5.5}$$

proved in Sewell (1973a, § IIc). The extrema are not in general stationary. A suffix zero refers to a solution value for the whole problem (5.1) + (5.2).

(iii) General bounds for linear functionals

In place of (2.9), choose

$$x_+ \text{ arbitrary, any } u_+ \geq 0, \text{ and } x_-, u_- = x_0, u_0 . \tag{5.6}$$

The consequent (2.1), when added to (5.4) to eliminate L_0, is

$$\begin{aligned} &L_+ - (x_+, \frac{\partial L}{\partial x_+}) - L_\beta + \langle u_\beta, \frac{\partial L}{\partial u_\beta}\rangle \\ &\geq -(x_0, \frac{\partial L}{\partial x_+}) + \langle u_0, \frac{\partial L}{\partial u_\beta}\rangle \\ &\geq -(x_0, \frac{\partial L}{\partial x_+}) . \end{aligned} \tag{5.7}$$

The left side is a supposedly known estimate for either of the two linear functionals of x_0 and u_0 on the right.

Instead of (5.6), modify (2.11) to choose

$$x_+, u_+ = x_0, u_0 \quad \text{and any} \quad x_- \geq 0, \quad \text{arbitrary} \quad u_- . \tag{5.7}$$

The consequent (2.1), when added to (5.3) to remove L_0, is

$$\begin{aligned} &L_\alpha - (x_\alpha, \frac{\partial L}{\partial x_\alpha}) - L_- + \langle u_-, \frac{\partial L}{\partial u_-} \rangle \\ &\geq - (x_0, \frac{\partial L}{\partial x_\alpha}) + \langle u_0, \frac{\partial L}{\partial u_-} \rangle \\ &\geq \langle u_0, \frac{\partial L}{\partial u_-} \rangle . \end{aligned} \tag{5.8}$$

Again the left side is a supposedly known estimate for either of the two linear functionals of x_0 and u_0 on the right.

The general bounds (5.7) and (5.8) are extensions of (2.13) and (2.14). Optimization of them is unexplored, but their brevity warrants their inclusion here, for completeness.

(iv) Embedding method

When an ab initio problem (1.19) contains an operator $N(x)$ which happens to be convex in some domain, then we can construct a functional (4.1) for which the left side of (2.1) is

$$\langle u_+, N(x_-) - N(x_+) - N'(x_+)(x_- - x_+) \rangle \tag{5.9}$$

which will be nonnegative in the half-space $u_+ \geq 0$, as in the case of (2.25).

Then (4.1) is a saddle functional. This suggests seeking to embed the problem in a variant of (5.1) and (5.2), namely

$$-N'^{*}(x)u + p = 0 \,, \qquad (\alpha) \qquad (5.10)$$

$$\left.\begin{aligned} -N(x) &\geq 0 \,, & (\beta) \\ u &\geq 0 \,, & (\alpha) \\ -\langle u, N(x)\rangle &= 0 \,. & \end{aligned}\right\} \qquad (5.11)$$

Thus we take (1.13α) with (5.2). The orthogonality condition is taken to imply that $N(x) = 0$ whenever the strict inequality $u > 0$ holds, and in that sense the embedding is achieved.

The objective now is therefore to bound (p, x_0) corresponding to $u_0 > 0$ in the actual solution of (5.10) with (5.11) (and not to bound (p, x_β) as in §4, because (5.11β) is not itself the ab initio problem). The dual extremum principles (5.5) still apply, and for (4.1) become

$$-\langle u_\alpha, N(x_\alpha)\rangle + (p, x_\alpha) \geq (p, x_0) \geq (p, x_\beta) \,. \qquad (5.12)$$

These are themselves the required bounds. The bound on the right is not necessarily stationary because possibly first order terms have been given away in its derivation, but it may be easy to find.

In the algebraic example (3.13), the ab initio problem was the quartic

$$N(x) \equiv \frac{1}{4}x^4 + \frac{1}{2}ax^2 + bx + c = 0 \,. \qquad (4.13\beta)$$

There are either one or two domains in which it is convex, and at most one domain for which it is concave (for which case the function can first be turned upside down before applying the procedure). In a convex domain

the bounds (5.12) read

$$-u_\alpha\left(\frac{1}{4}x_\alpha^4 + \frac{1}{2}ax_\alpha^2 + bx_\alpha + c\right) + px_\alpha \geq px_0 \geq px_\beta \tag{5.13}$$

where x_β is any solution of

$$\frac{1}{4}x_\beta^4 + \frac{1}{2}ax_\beta^2 + bx_\beta + c \leq 0 \tag{5.14}$$

and x_α is anything for which

$$(x_\alpha^3 + ax_\alpha + b)u_\alpha = p, \quad u_\alpha \geq 0 \tag{5.15}$$

can be satisfied. Evidently the cubic coefficient ought not to vanish in (5.15), and (4.9) is a formal way of avoiding this.

The embedding of ab initio linear equations can also be illustrated, either via a form of (4.10), or by embedding in a linear programming problem (cf. Noble and Sewell, op. cit. §10(ii)). Linear problems in which the operator has special structure have been discussed by Barnsley and Robinson (1974, 1975/6).

6. Applications

(i) Introduction

We have carried out some preliminary calculations applying the *general optimization method* of § 2. These include an analysis of an electrical network with resistors having nonlinear voltage-current relationships, and a verification that a basis used by Martin (1964, inequality (21)) for displacement bounds in elastic bodies under certain dynamic conditions is a consequence of ideas like those of (2.13) or (2.14) above. Barnsley and Robinson (1976) illustrate the result (4.12) by applications to a non-linear integral equation in communication theory, and a nonlinear differential equation in a thermal problem. Fujita (op. cit.) mentions examples for the linear problem (1.1).

We have concluded, however, that a fully representative illustration of the optimization method merits a separate investigation which we ought not to attempt here. A comparative study of the relative merits and power of the three methods described in §§ 2, 4 and 5 must also await the study of a number of examples.

The main objective of the present paper has been to establish some perspective by trying to uncover the structure of the requisite general theories. One may anticipate that in some later instances more rigorous statements may be required, but we have not conceived that to be necessary for our purpose here.

(ii) Nonlinear cantilever beam

In particular it is by no means clear from the literature that one-dimensional problems are genuinely representative of a theory which is to estimate pointwise bounds. Nevertheless it is a natural engineering starting point, and we conclude the paper by giving the reader a handle to the machinery of § 2 in such a case. This example was examined by Martin (1966) by an ad hoc engineering analysis, and a description from scratch of some of its connections with the present theoretical framework was given by Noble (1974) at an earlier stage of this research.

(iii) Hamiltonian representation of the beam problem

We first show how the elementary governing equations of the problem can be cast into the Hamiltonian form (1.16). This will illustrate how the appropriate spaces and operators can be constructed ab initio in a one-dimensional problem. Corresponding material in three-dimensional boundary value problems of elasticity and plasticity was given by Sewell (1973a,b).

We consider a thin straight cantilever beam made of nonlinear material. After conversion to nondimensional variables, let s denote distance measured along the beam from the built-in end $s = 0$ to the free end $s = 1$. Suppose the beam is loaded transversely in a plane by a load $w(s)$ per unit length. (See Fig. 6.1). The transverse small deflection (or deflection-'rate') in the direction of $w(s)$ is denoted by $u(s)$, and $M(s)$ is the

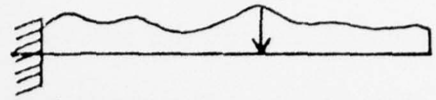


Fig. 6.1. Continuously loaded thin cantilever

internal bending moment. With an appropriate sign convention, elimination of the transverse internal shear force by differentiation leads to the single equilibrium equation

$$\frac{d^2M}{ds^2} = w(s) . \tag{6.1}$$

The boundary conditions will be

$$u(0) = \left.\frac{du}{ds}\right|_0 = 0, \qquad M(1) = \left.\frac{dM}{ds}\right|_1 = 0 . \tag{6.2}$$

The material is supposed to respond according to the 'creep law'

$$\frac{d^2u}{ds^2} = M^n \tag{6.3}$$

for some given n.

Our purpose in this sub-section is to express (6.1) - (6.3) in the formalism of (1.16). The space E is chosen to consist of matrices like

$$M \equiv \begin{bmatrix} M(s) \\ M(0) \\ 0 \end{bmatrix} \tag{6.4}$$

constructed from real integrable functions $M(s)$, the three entries being associated respectively with the interior and with the two end-points $s = 0$ and $s = 1$ of the beam. (The identity symbol emphasizes a definition.) The inner product for E is defined as

$$(M, N) \equiv \int_0^1 M(s)N(s)ds + M(0)N(0) \tag{6.5}$$

for any two members M and N of E. The space F is chosen to consist of matrices like

$$u \equiv \begin{bmatrix} u(s) \\ 0 \\ u(1) \end{bmatrix} \tag{6.6}$$

constructed from real integrable functions $u(s)$, the three entries again being associated with the interior and the end-points of $0 \leq s \leq 1$, with the same ordering as in (6.4). The inner product for F is defined as

$$\langle u, v \rangle \equiv \int_0^1 u(s)v(s)ds + u(1)v(1) \ , \tag{6.7}$$

for any two members u and v of F. Notice that there is a slight clash between the notation M, u just introduced by the definitions (6.4) and (6.6) for elements of the spaces, and the conventional way in which the real scalar <u>functions</u> $M(s)$, $u(s)$ have been abbreviated in (6.1) - (6.3) by omitting explicit mention of the argument s. This need not cause confusion.

It would have been possible to redefine E and F, by replacing the zero value entries in (6.4) and (6.6) by the values $M(1)$ and $u(0)$ (respectively) of the considered integrable functions, regarding these values as unassigned at this stage (they would <u>later</u> be given zero values in the subspaces E' and F'). Then $M(1)N(1)$ could have been added to the definition of (M, N), and $u(0)v(0)$ to that of $\langle u, v\rangle$. The two inner product spaces would <u>then</u> in fact be the same space. But there is no advantage in that, for we shall next be obliged to consider subspaces E' and F' which are <u>not</u> the same. In any event, from the viewpoint of general theory it is more fruitful to regard the presence of two (occasionally more) distinct spaces as the rule, and their coincidence as an exception.

The subspace E' is now defined to consist of those elements (6.4) of E which are constructed from functions (typically $M(s)$) which are not merely integrable, but <u>also</u>

are single-valued and continuous, with continuous first derivatives,

in $0 \leq s \leq 1$;

have piecewise continuous second derivatives in $0 < s < 1$;

have zero values at $s = 1$, e.g. $M(1) = 0$.

The subspace F' is defined to consist of those elements (6.6) of F which are constructed from functions (typically $u(s)$) which again are not merely integrable, but also

are single-valued and continuous, with continuous first derivatives,

in $0 \le s \le 1$;

have piecewise continuous second derivatives in $0 < s < 1$;

have zero values at $s = 0$, e.g. $u(0) = 0$.

The last property in each of these definitions shows that $E' \neq F'$, even though we could have chosen $E = F$ as described above.

We can now define operators T and T^* mapping according to (1.3) by the matrices

$$T^* u \equiv \begin{bmatrix} \dfrac{d^2u}{ds^2} \\ \left.\dfrac{du}{ds}\right|_0 \\ 0 \end{bmatrix} , \tag{6.8}$$

$$TM \equiv \begin{bmatrix} \dfrac{d^2M}{ds^2} \\ 0 \\ -\left.\dfrac{dM}{ds}\right|_1 \end{bmatrix} . \tag{6.9}$$

It can be verified that the statement (1.4) of adjointness, namely

$$(M, T^* u) = \langle u, TM \rangle \tag{6.10}$$

for all u in F' and for all M in E', here represents a double integration by parts written as

$$\int_0^1 M \frac{d^2u}{ds^2} ds + M(0) \left.\frac{du}{ds}\right|_0 = \int_0^1 u \frac{d^2M}{ds^2} ds - u(1) \left.\frac{dM}{ds}\right|_1 . \tag{6.11}$$

The jumps allowed in the second derivatives do not affect the validity of this, and the properties $M(1) = 0 = u(0)$ of the subspaces have been used.

Finally we can introduce the Hamiltonian functional

$$X[M,u] = \int_0^1 \left[\frac{1}{n+1} M^{n+1} + uw\right] ds \; . \tag{6.12}$$

This has no boundary terms, which is exceptional, and so its gradients are

$$\frac{\partial X}{\partial M} = \begin{bmatrix} M^n \\ 0 \\ 0 \end{bmatrix} , \tag{6.13}$$

$$\frac{\partial X}{\partial u} = \begin{bmatrix} w \\ 0 \\ 0 \end{bmatrix} . \tag{6.14}$$

The equations (1.16) now appear as

$$\begin{aligned} T^* u &= \frac{\partial X}{\partial M} \qquad (\alpha) , \\ T M &= \frac{\partial X}{\partial u} \qquad (\beta) \end{aligned} \tag{6.15}$$

which can be seen as an alternative statement of the original equations (6.1) - (6.3), bearing in mind also the properties of E' and F'.

The problem is thus generated from equations (6.15) by the Hamiltonian functional $X[M,u]$ of (6.12), which is strictly convex in M if n is an odd integer (or convex in the half-space $M \geq 0$ if n is any integer), and linear in u. A classical elastic beam has $n = 1$.

Thus the problem is very similar to the Fujita problem (1.14) when $n = 1$, or its generalization (2.26) when $n > 1$, except that the role of the variables is reversed. In applying the general theory, we therefore expect to have a case of intermediate generality like that of subsections 2(vi) - (viii).

(iv) Lagrangian generating functional

Evidently (6.15α) can be regarded as the 'constitutive equation', and (6.15β) as the 'equilibrium equation'. The quote marks remind us that these equations in fact contain some of the boundary conditions embedded in them as well. The equations can also be regarded as generated from (1.13) via the Lagrangian functional

$$\begin{aligned} L[M,u] &= (M, T^* u) - \int_0^1 \left[\frac{1}{n+1} M^{n+1} + uw\right] ds \\ &= \int_0^1 M \frac{d^2u}{ds^2} ds + M(0) \left.\frac{du}{ds}\right|_0 - X[M,u] \\ &= \langle u, TM \rangle - X[M,u] \\ &= \int_0^1 u \frac{d^2M}{ds^2} ds - u(1) \left.\frac{dM}{ds}\right|_1 - X[M,u] \; . \end{aligned} \tag{6.16}$$

In other words, the constitutive equations may be derived as

$$\frac{\partial L}{\partial M} = T^* u - \frac{\partial X}{\partial M} = \begin{bmatrix} \dfrac{d^2u}{ds^2} - M^n \\ \left.\dfrac{du}{ds}\right|_0 \\ 0 \end{bmatrix} = 0 \, , \tag{6.17α}$$

and the equilibrium equations as

$$\frac{\partial L}{\partial u} = TM - \frac{\partial X}{\partial u} = \begin{bmatrix} \dfrac{d^2M}{ds^2} - w \\ 0 \\ -\left.\dfrac{dM}{ds}\right|_1 \end{bmatrix} = 0 . \tag{6.17β}$$

The underdetermined class of solutions M_α, u_α of (6.17α), and M_β, u_β of (6.17β), are generated from (6.4) and (6.6) as follows. The element u_α must be constructed from a function $u_\alpha(s)$ which satisfies $u_\alpha(0) = 0$, because it must belong to the domain F' of T^*. By a double integration of (6.17α) with any integrable function $M_\alpha(s)$, using dummy variables σ and t, we have

$$u_\alpha(s) = \int_0^s \int_0^t [M_\alpha(\sigma)]^n d\sigma dt . \tag{6.18}$$

The element M_β must be constructed from a function $M_\beta(s)$ which satisfies $M_\beta(1) = 0$, because it must belong to the domain E' of T. By a double integration of (6.17β) with any integrable loading function $w(s)$, we have

$$\begin{aligned} M_\beta(s) &= \int_1^s \int_1^t w(\sigma) d\sigma dt \\ &= \frac{1}{2} w(s-1)^2 \quad \text{if} \quad w(\sigma) = \text{constant} . \end{aligned} \tag{6.19}$$

Nothing need be said about a function $u_\beta(s)$ or its associated element u_β, because this is absent from (6.17β).

The total potential energy associated with any such α-solution is

$$L_\alpha = (M_\alpha, \frac{\partial X}{\partial M_\alpha}) - X[M_\alpha, u_\alpha]$$

$$= \int_0^1 \left[\frac{n}{n+1} M_\alpha^{n+1} - u_\alpha w\right] ds \tag{6.20}$$

$$= \int_0^1 \left[\frac{n}{n+1} M_\alpha^{n+1} - M_\beta M_\alpha^n\right] ds \ .$$

The total complementary energy associated with any such β-solution is

$$-L_\beta = -\langle u_\beta, \frac{\partial X}{\partial u_\beta}\rangle + X[M_\beta, u_\beta]$$

$$= \int_0^1 \frac{1}{n+1} M_\beta^{n+1} ds \ . \tag{6.21}$$

Any actual solution value L_0 of L is

$$L_0 = \int_0^1 \left[\frac{n}{n+1} M_0^{n+1} - u_0 w\right] ds = -\int_0^1 \frac{1}{n+1} M_0^{n+1} ds$$

$$= -\int_0^1 u_0 \frac{w}{n+1} ds \ . \tag{6.22}$$

In other words, L_0 is itself a linear functional of u_0.

When n is odd, $L[M, u]$ is a saddle functional concave in M and linear in u, and the standard energy methods are applications of the extremum principles (2.5) with these specific expressions (6.18) - (6.21). The difference between the energy bounds for odd n is

$$L_\alpha - L_\beta = \frac{1}{n+1}\int_0^1 [\,nM_\alpha^n(M_\alpha - M_\beta) - M_\beta(M_\alpha^n - M_\beta^n)]\,ds \geq 0$$

$$= \frac{1}{2}\int_0^1 (M_\alpha - M_\beta)^2 ds \quad \text{when} \quad n = 1 . \tag{6.22}$$

(v) A strong bound

Recalling the remarks of § 2(x) and noticing that (6.22) reduces L_0 itself to a linear functional of u_0, we enquire if a strong bound can be constructed for the beam problem.

Substitution of the Lagrangian (6.16) into (2.12) leads to

$$\frac{1}{n+1}\int_0^1 M_-^{n+1} ds \geq \int_0^1 u_0\left[\frac{d^2M_-}{ds^2} - \frac{nw}{n+1}\right]ds - u_0(1)\left.\frac{dM_-}{ds}\right|_1 \tag{6.23}$$

after using the matrix expression (6.17β) for the nonzero gradient $\partial L/\partial u_-$ in the inner products defined via (6.7).

Since the minus point in (2.12) is arbitrary, so is the bending moment distribution $M_-(s)$ except that it must be in E', the domain of T specified above for this problem.

This makes a precise connection with Martin's (1966) result, if we now choose $M_-(s)$ to be in equilibrium with the fictitious loading distribution shown in Fig. 6.2. That is to say, $M_-(s)$ is to satisfy

$$\frac{d^2M_-}{ds^2} = \frac{nw(s)}{n+1} \quad \text{in} \quad 0 < s < 1 \tag{6.24}$$

$$-\frac{dM_-}{ds} = P \quad \text{at} \quad s = 1 \tag{6.25}$$

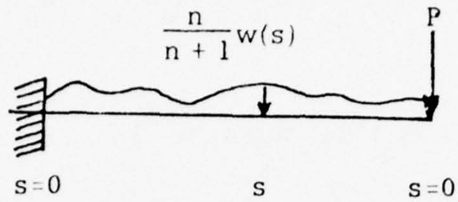


Fig. 6.2. Fictitious loading on cantilever

in addition to $M_-(1) = 0$ already required by the subspace E', where P is a given number. Such a choice is made because it allows the <u>pointwise estimate</u>

$$u_0(1) \leq \frac{1}{(n+1)P} \int_0^1 M_-^{n+1} ds \tag{6.26}$$

to be obtained from (6.23) for the deflection $u_0(1)$ at the end of the beam under the <u>actual</u> loading of Fig. 1.

When the distributed load w is uniform, the required solution of (6.24) is

$$M_-(s) = \frac{n}{n+1} \frac{1}{2} w(1 - s)^2 + P(1 - s) , \tag{6.27}$$

whence (6.26) becomes Martin's (1966) pointwise estimate (12).

<u>(vi) Weak bounds</u>

It is possible, for example, to optimize (2.14) by identifying the above choice of M_- with

$$M_- = M_\beta + hp \tag{6.28}$$

where h is a scalar and p a member of E'. But we already know that, in a sense, more than necessary has been given away in this particular problem, and the result does not seem to be helpful. For example, in the case $n = 1$ we arrive at

$$\left|\int_0^1 (M_0 - M_\beta)p\,ds\right| \leq \left[\int_0^1 (M_\alpha - M_\beta)^2 ds\right]^{\frac{1}{2}} \left[\int_0^1 p^2 ds\right]^{\frac{1}{2}} . \tag{6.29}$$

A reason why we said that this problem has only limited representative value can now be seen. It is because, since we are at liberty to choose any $M_\alpha(s)$ for insertion into (6.18), and since we know $M_\beta(s)$ from (6.19), we can choose

$$M_\alpha(s) = M_\beta(s) . \tag{6.30}$$

This is the <u>perfect</u> choice bearing no margin of error in (6.29), and in fact corresponds to the <u>exact</u> solution $M_0(s)$. The exact solution when $n = 1$ and $w = \text{constant}$ is

$$\begin{aligned} M_0(s) &= \frac{1}{2} w(s - 1)^2 \\ u_0(s) &= \frac{1}{24} ws^2(s^2 - 4s + 6) . \end{aligned} \tag{6.31}$$

REFERENCES

Barnsley, M. F. and Robinson, P. D. 1974 Bivariational bounds, Proc. R. Soc. Lond. A. 338, 527-533.

1975/6 Bivariational bounds associated with non-self-adjoint linear operators, Proc. R. Soc. Edin. 75A (9), 109-118.

1976 Bivariational bounds for nonlinear problems, to be published.

Cartan, H. 1971 Differential Calculus, Herman, Paris.

Fujita, H. 1955 Contribution to the theory of upper and lower bounds in boundary value problems, J. Phys. Soc. Japan, 10, 1-8.

Hill, R. 1956 New horizons in the mechanics of solids, J. Mech. Phys. Solids 5, 66-74.

Kato, T. 1953 On some approximate methods concerning the operators T^*T, Math. Annalen 126, 253-264.

Martin, J. B. 1964 A displacement bound technique for elastic continua subjected to a certain class of dynamic loading, J. Mech. Phys. Solids 12, 165-175.

1966 A note on the determination of an upper bound on displacement rates for steady creep problems, J. Appl. Mech. 33, 216-217.

Noble, B. 1964 Complementary variational principles for boundary value problems I: Basic principles with an application to ordinary differential equations, Mathematics Research Center Report 473, University of Wisconsin, 30 pp.

1974 Pointwise bounds in nor'inear problems, Bulletin of the Institute of Mathematics and its Applications 10, 103-107.

Noble, B. and Sewell, M. J. 1972 On dual extremum principles in applied mathematics, J. Inst. Maths. Applics. 9, 123-193.

Rall, L. B. 1969 Computational Solution of Non-linear Operator Equations, Wiley, New York.

Sewell, M. J. 1969 On dual approximation principles and optimization in continuum mechanics, Phil. Trans. Roy. Soc. Lond. A265, 319-351.

1973 The governing equations and extremum principles of elasticity and plasticity generated from a single functional,

a J. Struct. Mech. 2(1), 1-32

b Ibid. 2(2), 135-158.

1976 Some mechanical examples of catastrophe theory, Bulletin of the Institute of Mathematics and its Applications 12 163-172.

Zago, J. V. 1976 Ph.D. Thesis, University of Wisconsin.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 1703	2. GOVT ACCESSION NO. 9	3. RECIPIENT'S CATALOG NUMBER Technical
4. TITLE (and Subtitle) 6 GENERAL ESTIMATES FOR LINEAR FUNCTIONALS IN NONLINEAR PROBLEMS.		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) 10 M. J. Sewell and B. Noble	15	8. CONTRACT OR GRANT NUMBER(s) DAAG29-75-C-0024
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS See Item 18 below.	11	12. REPORT DATE December 1976
		13. NUMBER OF PAGES 64
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	12 68p.	15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

14 MRC-TSR-1703

18. SUPPLEMENTARY NOTES

U. S. Army Research Office	University of Oxford	University of Reading
P. O. Box 12211	Computing Laboratory	Department of Mathematics
Research Triangle Park	Oxford, England	United Kingdom
North Carolina 27709		

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Approximation of linear functionals
Optimization, embedding, nonlinear programming
Extremum principles, variational principles
Pointwise bounds
Saddle functional

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A general theory for estimating linear functionals in nonlinear problems is given, in terms of the gradients of a generating Lagrangian functional. The theory is illustrated by finding bounds on the deflection of a nonlinear cantilever beam.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

221200

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

1/B

DAT

ILM